# 14 Structure, Dynamics, Activity, and Function of Compstatin and Design of More Potent Analogues

*Dimitrios Morikis and John D. Lambris*

## CONTENTS



## I. INTRODUCTION

The activation of complement system is a complex process finely controlled by regulators of complement activation (RCA) proteins or natural inhibitors. This type of regulation is important to direct complement function against invading foreign

pathogens and waste products of immune system reactions and to enhance antibody responses (reviewed in References 1 through 4). In addition, components of the complement system are used by viruses to enter host cells (reviewed in References 1 and 5) and recent studies have shown that components of the complement system are involved in developmental processes such as organ regeneration and hematopoietic development.[5,6] When the complement system is inappropriately activated or when its regulation breaks down, it is capable of attacking normal tissues with harmful effects to the host. Similarly, a hereditary deficiency of a complement component is usually responsible for breakdown of the beneficial effects of complement activation for immune response (reviewed in References 1 and 2). Various studies have shown that the complement system has been involved in a number of pathological situations, including autoimmune diseases, degenerative diseases, ischemia/reperfusion injuries, burn injuries, asthma, hemodialysis, cardiopulmonary bypass surgery, and transplantation (reviewed or compiled in lists in References 1 and 7 through 10). Currently there are no clinically available anticomplement drugs, despite intense research and numerous candidates, some of which have made it to clinical trials. Complement activation inhibitors range from natural inhibitors and regulators, monoclonal antibodies, peptides, natural products, and organic molecules (reviewed in References 8 through 12).

Complement component C3 (reviewed in References 13 and 14) is an excellent target for inhibition because it is the convergence point of the classical, lectin, and alternative pathways of complement activation and the starting point of the common pathway. In this chapter, we present an overview of the discovery of the C3-binding complement inhibitor peptide compstatin and the design of active compstatin analogues. The latter will be called hereafter "active analogues" or "higher inhibitory activity analogues," when this is the case compared to parent peptide compstatin. This work spans about 9 years of research using a diverse set of tools from the fields of immunology, molecular biology, spectroscopy, structural, biology, protein and peptide chemistry, biochemistry, computational chemistry, biophysics, combinatorial and global optimization, and bioengineering. The design of active compstatin analogues is an example of a methodologically integrative, cross-disciplinary, and collaborative approach. Reviews on the discovery and design of compstatin and analogues at various stages of the process have been published[15,16] including comparisons to other complement inhibitors[9,17,18] and in view of complement research in general.[5]

## II. DISCOVERY OF COMPSTATIN USING A PHAGE-DISPLAYED RANDOM PEPTIDE LIBRARY

Random peptide libraries displayed in phages have proven to be useful tools for large and rapid combinatorial searches for peptides that bind to proteins or monoclonal antibodies (reviewed in Reference 19). Compstatin is a truncated derivative of a peptide that was discovered by means of screening a phage-displayed random peptide library for binding to C3b by Sahu et al.[20] The phage-displayed random peptide library contained $2 \times 10^8$ unique clones expressing 27-residue random

peptides with sequence Ser-Arg-$Xaa_{12}$-(Ser,Pro,Thr,Ala)-Ala-(Val,Ala,Asp, Glu,Gly)-$Xaa_{12}$-Ser-Arg, where $Xaa_{12}$ is a 12-member sequence of any of the 20 natural amino acids separated by one fixed and two semifixed amino acids. An active clone isolated from the phage exhibited specific binding to immobilized C3, C3b, and C3c, but not to C3d.[20] Similar results were obtained in an ELISA assay using a 27-residue synthetic peptide corresponding to the sequence of the C3-binding clone. Also, the synthetic peptide inhibited both the alternative and the classical pathways of complement activation in hemolytic assays of rabbit and antibody-coated sheep erythrocytes, respectively, in normal human serum.[20] This inhibition was shown to be reversible using gel filtration experiments. The sequence of the 27-residue synthetic peptide was Ile-[Cys-Val-Val-Gln-Asp-Trp-Gly-His-His-Arg-Cys]-Thr-Ala-Gly-His-Met-Ala-Asn-Leu-Thr-Ser-His-Ala-Ser-Ala-Ile-$NH_2$ (hereafter, brackets around cysteines denote cyclization through disulfide bonds). Sequence truncation demonstrated that the minimum length peptides that showed complement inhibitory activities had sequences Ile-[Cys-Val-Val-Gln-Asp-Trp-Gly-His-His-Arg-Cys]-Thr-$NH_2$[20] and [Cys-Val-Val-Gln-Asp-Trp-Gly-His-His-Arg-Cys]-$NH_2$.[21] The former had about threefold higher inhibitory activity than the latter (Table 14.1), and was named compstatin.[21] Compstatin also bound to C3, C3b, C3c, but not to C3d.[22] Cyclization was found to be important for activity, as linear peptides produced by reduction and alkylation of cysteines[20,22] or by replacing cysteines with alanines[22,23] were inactive. Another systematic search for a shorter active analogue by stepwise deletion of residues within the cyclization loop was not successful.[22] It should be noted that compstatin inhibited the alternative pathway at about fivefold lower concentration compared to the classical pathway, perhaps because the alternative pathway is more sensitive to activation and deposition of C3 to target surfaces.[20]

## III. BIOLOGICAL MECHANISM OF INHIBITION

The inhibition of complement activation by compstatin was measured in normal human serum and found to be almost identical to that of the alternative pathway using hemolytic assays.[20] It was shown that complement inhibition was the result of inhibition of the cleavage of C3 to C3a and C3b by C3 convertase enzymes.[20,22] However, the composition of the assay and the complexity of the cascade of complement activation raised questions on the possibility of additional mechanisms of inhibition. Control experiments, some of which were indirect, led to the following hypotheses:[20]

1. Compstatin did not inhibit the cleavage (and spontaneous inactivation) of C3b to iC3b mediated by factors H and I.[20]
2. Compstatin did not inhibit the association of C3b with factor Bb to form the C3 convertase enzyme C3bBb. It was shown that compstatin did not affect the cleavage of factor B to Ba and Bb by factor D, thus allowing the formation of the C3 convertase enzyme complex by C3b and Bb.[20] But, since compstatin also binds to C3b, it was not clear that it did not have an effect on the catalytic function of the convertase enzyme. Another

**TABLE 14.1**
**Selected Active Analogues of Compstatin Discussed in Text**

| # | Peptide[a] | Sequence[b] | Relative Activities[c] | Reference |
|---|---|---|---|---|
| ***1*** | ***Compstatin ring*** | ***[CVVQDWGHHRC]-NH$_2$*** | ***0.4*** | ***21*** |
| 2 | Compstatin ring/$C_{ter}$-flanking | [CVVQDWGHHRC]T-NH$_2$ | 0.5 | 22 |
| 3 | Ac-I1S/V4F/H9R/H10L/R11A/T13P | **Ac-S**[CV**F**QDWG**RLA**C]**P**-NH$_2$ | 0.5 | 25 |
| 4 | Ac-R11S | **Ac**-I[CVVQDWGHH**S**C]T-NH$_2$ | 0.5 | 23 |
| 5 | Ac-I1D | **Ac-D**[CVVQDWGHHRC]T-NH$_2$ | 0.5 | 25 |
| 6 | R11K | I[CVVQDWGHH**K**C]T-NH$_2$ | 0.6 | 23 |
| 7 | Compstatin ring/H9A | [CVVQDWGAHRC]-NH$_2$ | 0.8 | 21 |
| ***8*** | ***Compstatin*** | ***I[CVVQDWGHHRC]T-NH$_2$*** | ***1*** | ***20*** |
| 9 | Ac-V3L | **Ac**-I[C**L**VQDWGHHRC]T-NH$_2$ | 1 | 23 |
| 10 | Ac-H9A/R11A | **Ac**-I[CVVQDWG**A**H**A**C]T-NH$_2$ | 1 | 23 |
| 11 | Ac-V3L/Q5N | **Ac**-I[C**L**V**N**DWGHHRC]T-NH$_2$ | 1 | 23 |
| 12 | Ac-I1R | **Ac-R**[CVVQDWGHHRC]T-NH$_2$ | 2 | 23 |
| ***13*** | ***Ac-compstatin*** | ***Ac-I[CVVQDWGHHRC]T-NH$_2$*** | ***3*** | ***22*** |
| 14 | Ac-Q5N | **Ac**-I[CVV**N**DWGHHRC]T-NH$_2$ | 3 | 23 |
| 15 | Ac-V4A/H9A/T13I | **Ac**-I[CV**A**QDWG**A**HRC]**I**-NH$_2$ | 3 | 23 |
| 16 | Ac-T13I | **Ac**-I[CVVQDWGHHRC]**I**-NH$_2$ | 4 | 23 |
| 17 | Ac-H9A | **Ac**-I[CVVQDWG**A**HRC]T-NH$_2$ | 4 | 23 |
| 18 | Ac-I1L/H9W/T13G | **Ac-L**[CVVQDWG**W**HRC]**G**-NH$_2$ | 4 | 25 |
| 19 | Ac-I1V/V4Y/H9F/T13V | **Ac-V**[CV**Y**QDWG**F**HRC]**V**-NH$_2$ | 6 | 40 |
| 20 | Ac-I1V/V4Y/H9A/T13V | **Ac-V**[CV**Y**QDWG**A**HRC]**V**-NH$_2$ | 9 | 40 |
| 21 | Ac-V4Y/H9F/T13V | **Ac**-I[CV**Y**QDWG**F**HRC]**V**-NH$_2$ | 11 | 40 |
| 22 | Ac-V4Y/H9A/T13V | **Ac**-I[CV**Y**QDWG**A**HRC]**V**-NH$_2$ | 14 | 40 |
| 23 | Ac-V4Y/H9A | **Ac**-I[CV**Y**QDWG**A**HRC]T-NH$_2$ | 16 | 40 |
| 24 | Ac-V4W/H9A | **Ac**-I[CV**W**QDWG**A**HRC]T-NH$_2$ | 45 | 49 |

[a] Analogues are arranged in order of increasing activity.

[b] Brackets denote cyclization through a disulfide bridge of Cys2–Cys12. One-letter amino acid code is used for simplicity. Compstatin, compstatin ring, and acetylated compstatin are in boldface italic. Amino acid replacements in each analogue compared to compstatin are in boldface.

[c] Approximate relative activities. Small deviations from published data are owed to subsequent repeat measurements using different assays.

study[24] measured the direct generation of the C3bBb convertase using surface plasmon resonance, in a step-wise manner by repeating factor B, factor D, and C3 reactions, followed by addition of compstatin, C3+compstatin, or C3+control peptide (control was linear inactive compstatin). This study showed no direct effect of compstatin on the C3-cleaving ability of the C3bBb convertase. However, addition of C3+compstatin resulted in no further increase in the formation of C3bBb in contrast to addition of C3+control peptide that allowed generation of more C3bBb. Conceptually, compstatin should have an indirect effect in the formation of C3 convertase, which needs the C3 cleavage product C3b. C3b becomes unavailable

for the convertase complex formation when compstatin binds to C3 and inhibits its cleavage. In this sense, the levels of the C3 convertase should remain unaltered. This expectation was experimentally confirmed by the surface plasmon resonance study of Nilsson et al.[24]

3. Compstatin did not sterically obstruct the access of the C3 convertase enzyme C3bBb to the C3a/C3b cleavage site at Arg726-Ser727.[20] It was shown that trypsin was capable of cleaving C3 in complex with an arginine-free active compstatin analogue, while the C3 convertase C3bBb failed to do so. The Arginine-free (Arg11Ala) analogue guaranteed activity (albeit lower than compstatin) and at the same time preserved the integrity of its backbone from cleavage by trypsin. It should be noted that this is not a direct comparison because C3 convertase and trypsin are enzymes of different size and shape.
4. Compstatin did not have a significant effect in the binding of properdin to C3, suggesting that compstatin did not disrupt the formation of properdin-stabilized C3 convertase, C3bBbP.[20]
5. The inhibition of complement activation in human blood, plasma, and serum was owed to C3 binding and not to proteolytic cleavage of compstatin resulting in a smaller or a linear peptide. Proteolytic cleavage is possible at $^{+}H_3N$-Ile1-Cys2 and Arg11-Cys12 peptide bonds, but not at the blocked (and neutral) carboxy-terminus Cys12-Thr13-$NH_2$ peptide bond. Compstatin showed limited proteolytic cleavage, which was slow at Ile1 (rates of 0.03% per minute at 37°C and 0.01% per minute at 22°C) and very slow at Arg11.[22,25] Other biotransformation products were not observed for compstatin as opposed to a linear analogue (with Cys to Ala replacements) that was efficiently processed starting from the amino-terminus.[22] This led to the hypothesis that the cyclic nature of compstatin prevents its enzymatic processing beyond Ile1.

The fact that compstatin inhibited 50% of the alternative pathway complement activity at a concentration only about twofold higher than the concentration of C3 in normal human serum makes it a promising candidate for potential therapeutic use.[20] The exact mechanism of inhibition is still unknown and possibly involves several subtleties. It is remarkable that compstatin exhibits inhibition of activation of complement from humans and primates (six monkey species were tested) but not from lower mammalian species.[26,20,27] Specifically, compstatin did not inhibit activation from pig, rabbit, guinea pig, rat, and mouse complement.[26] Also, in surface plasmon resonance experiments, compstatin did not exhibit binding to human C4 and C5, which are sequence and structural homologues of C3.[26] It was shown that the lack of mouse complement inhibition by compstatin is owed to differences in mouse and human C3, rather than to differences in other complement components involved in activation.[26] Experiments with reconstituted human C3 in C3-deficient mouse serum restored the inhibitory activity of compstatin, supporting the observation that it is specific to human (and primate) C3.[26] It is possible that there are structural differences between human and mouse C3 despite the high degree of sequence similarity of their C3. The specificity of compstatin requires nonhuman

primates or transgenic mice expressing human C3 for the testing of its *in vivo* efficacy.

Attempts to identify sequence similarity between compstatin and a fragment of another known protein have failed.[16,20] The longest sequence segment of compstatin that matched a sequence segment of another protein in a BLAST[28] database search was Val-Val-Gln-Asp-Trp-Gly.[16] This segment is part of the active site cavity of haloalkane dehalogenase and we believe that the similarity is fortuitous.

## IV. COMPSTATIN APPLICATIONS

Compstatin has been shown to inhibit *in vitro* complement activation in human blood, plasma, and serum.[20,22] It has also been shown that compstatin is stable (from enzymatic processing) in human blood, plasma, and serum. The acetylated form of compstatin (see below) had half time of 24 hours at 37°C in human blood.[22] In addition, compstatin has been tested to be effective in several models of potential clinical relevance. Compstatin has shown promise to control complement activation during or after cardiac surgery and cardiopulmonary bypass. Activation of the complement system occurs because of the formation of heparine/protamine complexes, bioincompatibility with plastic material in extracorporeal blood circulation, and because of reperfusion of ischemic tissues with blood. An *in vivo* study using baboons showed that compstatin inhibited heparine/protamine-induced complement activation without adverse effects in heart rate or systemic arterial, central venous, and pulmonary arterial pressures in a model of cardiac and postcardiac surgery.[29] In another study, compstatin was successfully tested in two models of extracorporeal contact of whole blood with polymer surfaces of tubing loops and microscope slides.[24] It was shown that compstatin inhibited the generation of C3a and soluble membrane attack complex and the binding of C3b and polymorphonuclear leukocytes in the polymer surface. It was also shown that blood cell counts were not affected in these studies. In another study, compstatin was successfully tested for complement inhibition by polymeric material.[30] In all cases of this study, compstatin was effective in reducing biomaterial-mediated neutrophil activation independent of polymer surface type.

Complement has been implicated in transplantation rejection, because uncontrolled complement activation can result in ischemia/reperfusion injury and postischemic organ damage (reviewed in Reference 31). Compstatin was successfully tested as a possible delay or prevention agent for hyperacute rejection in an *ex vivo* model of xenotransplantation.[32,33] It was shown that compstatin prolonged the survival of porcine kidneys perfused with human blood. Comparison with control experiments did not show significant differences in leukocyte activation, platelet counts, thrombospondin, soluble P-selectin, and β-thromboglobulin.[32,33]

In other applications compstatin blocked *Escherichia coli*–induced oxidative burst of granulocytes and monocytes, complement activation, and secretion of cytokine IL-8 in human whole blood.[34] Also, compstatin inhibited C-reactive, protein-induced complement activation on cell line SH-SY5Y.[35]

An AstraZeneca Pharmaceuticals study[27] confirmed the activity of compstatin and several of the results were reported in Sahu et al.[20,22] and Morikis et al.[21,23] This

study examined several analogues of compstatin, some with acetylation at the amino-terminus and some with D-amino acids. This study also presented an additional proof that complement activation inhibition by compstatin was because of its effect on C3 and not because of an effect on the alternative pathway C3 convertase C3bBb.[27] This was accomplished using an alternative assay with a C3 convertase consisting of a complex of cobra venum factor (CVF) with factor Bb, CVFBb. Complement inhibition using this assay was comparable to standard hemolytic immunoassay with C3bBb convertase complex. The AstraZeneca study also reported that compstatin was not significantly cytotoxic, and showed little or no inhibition of clotting or inhibition of key enzymes in the clotting cascade. Finally, this work reported a limited species specificity study for compstatin.[27]

## V. STRUCTURE AND DYNAMICS OF COMPSTATIN

Structural studies of compstatin and analogues were performed using nuclear magnetic resonance (NMR) spectroscopy.[21,23,25,49] NMR is efficient for data collection of numerous compounds or peptide analogues in solution that are generated during the process of a drug design project. The flexibility of small peptides in solution is challenging for crystallization and amenability to x-ray crystallographic studies. NMR does not have this limitation, as it can study molecules in solution. Indeed, NMR not only can be used to determine the three-dimensional structure, but it is also the experimental method of choice to explore the dynamic nature of peptides in solution. As opposed to general practices that focus on the use of static three-dimensional structures and structure–activity correlations, we believe that dynamics play a key role in function or activity. This is more obvious in the case of small ligands, peptides, drugs, and small substrates in solution, which are highly flexible and span complex conformational spaces. Our NMR studies are complemented or extended (certain properties can even be predicted) by the use of computational methods, such as energy minimizations, molecular dynamics simulations, and electrostatic calculations. Indeed, computational methods are an integral part for the determination, and/or refinement, of three-dimensional structures by NMR or x-ray crystallography. Furthermore, these structures can be used as starting points to theoretically study local or global motions that are not easily accessible by experiment. We should note that the high efficiency of NMR is limited to small peptides free in solution. This efficiency breaks down in the study of protein–peptide complexes for which typically cocrystallization is necessary for crystallographic studies, or elaborate NMR schemes are necessary when the overall molecular mass of the complex is not prohibiting for NMR studies.

The structure of a major conformer of compstatin was determined using two-dimensional NMR spectroscopy by Morikis et al.[21] and is deposited in the Protein Data Bank (PDB)[36] with code 1A1P. NMR observables, such as chemical shifts, $^{3}J_{HN\text{-}H\alpha}$ coupling constants, and nuclear Overhauser effects (NOEs) suggested conformational averaging[21] (discussed in text and supplementary material of Reference 21). This was not unusual since it is well known that small peptides in solution form ensembles of interconverting conformers. In certain instances, particularly for cyclic peptides, a conformation with high population is detectable by NMR and amenable

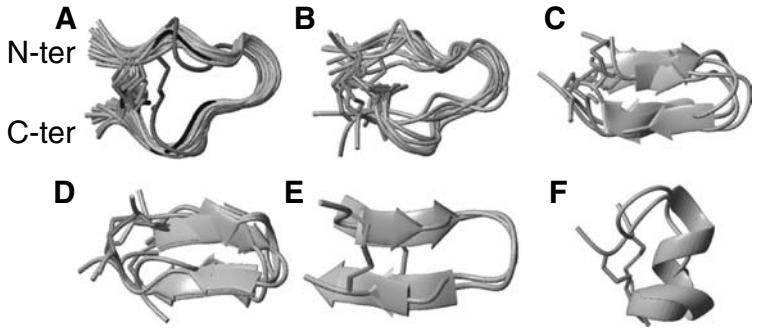


**FIGURE 14.1** Ribbon representations of the families of compstatin structures derived from nuclear magnetic resonance (NMR) spectroscopy with hybrid distance geometry/simulated annealing methodology,[21] global optimization methodology,[37] and molecular dynamics simulations.[38] The structures in each family are superimposed using the backbone heavy atoms, N, $C^{\alpha}$, C′, between residues 2 and 12. (A) Superimposition of the family of 21 NMR structures (grey), the average minimized NMR structure (black), and the global optimization structure (dark grey). The conformation is coil with type I β-turn (23 models in total). (B–F) Superimpositions of the families of five interconverting conformers of compstatin determined after 1 ns of molecular dynamics simulations. The following conformations were identified: (B) coil with type I β-turn (10 models), (C) β-hairpin with type I β-turn (4 models), (D) β-hairpin with type II′ β-turn (5 models), (E) β-hairpin with type VIII β-turn (2 models), and (F) coil with α-helix (2 models). Molecular graphics for this and subsequent figures were made with computer software MOLMOL.[51] Coordinates from structures with PDB code 1A1P (NMR ensemble) and companion CD codes ACST (average minimized NMR structure), CST_GOS (global optimization structure), and COMS_MD44, COMS_MD22, COMS_MD17, COMS_MD9A, and COMS_MD9B (molecular dynamics structures).

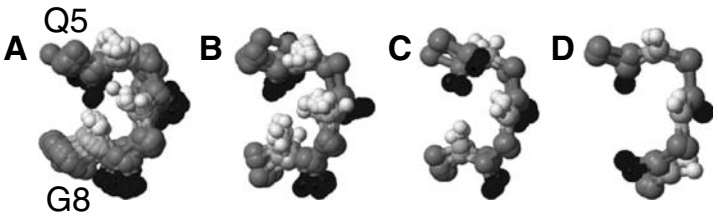


**FIGURE 14.2** Superimpositions of the β-turn segment, Gln5-Asp6-Trp7-Gly8, of the families of compstatin structures determined by nuclear magnetic resonance (NMR) spectroscopy[21,37] and molecular dynamics (MD) simulations.[38] (A) NMR/type I β-turn (23 models including the average minimized and global optimization structures). (B) MD/type I β-turn (14 models from both coil and β-hairpin families). (C) MD/type II′ β-turn (5 models). (D) MD/type VIII β-turn (2 models). The following color atom code has been used: black for oxygen, light grey for nitrogen, grey for carbon, and white for hydrogen. Molecular graphics and coordinates as in Figure 14.1.

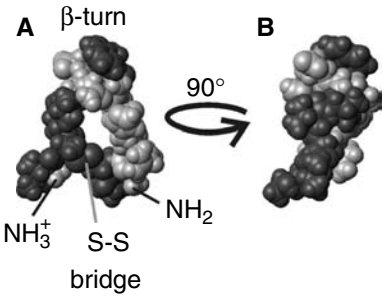


**FIGURE 14.3** Space-filling representation of the lowest energy structure of compstatin from nuclear magnetic resonance data.[21] The structural orientations in panels (A) and (B) are related by a 90° rotation as shown. Residues with hydrophobic character (Ile1, Cys2, Val3, Val4, Trp7, Cys12, Thr13) are drawn in black, and polar residues are drawn in grey (including amino-terminal group $NH_3^+$ and carboxy-terminal blocking group $NH_2$). The amino-terminal $NH_3^+$ group disrupts the hydrophobic cluster (see text). (Coordinates from PDB code 1A1P.)

to structure determination. This was the case of compstatin. A family of 21 low energy NMR structures was calculated using hybrid molecular dynamics/simulated annealing computational methodology and NMR-derived restraints. The structure of compstatin was also calculated using a novel global optimization methodology and a subset of the NMR-derived restraints.[37] Figure 14.1A shows a superimposition of the family of 21 NMR structures of compstatin, the average minimized NMR structure, and the global optimization structure. Compstatin forms a coil conformation with a type I β-turn (Figures 14.1A and 14.2A). The β-turn spans residues Gln5-Asp6-Trp7-Gly8 located opposite to the disulfide bridge of Cys2-Cys12 (Figures 14.1A, 14.2A). The disulfide bridge is part of a hydrophobic cluster at the surface and limited core of compstatin, spanning residues Ile1-Cys2-Val3-Val4/Cys12-Thr13, and the β-turn is part of the remaining polar part. Figure 14.3 shows the relative topologies of the hydrophobic and polar residues, and the β-turn and disulfide bridge. Figure 14.4 shows a map of the electrostatic potential on the surface of compstatin. The calculation of the electrostatic potential of compstatin involved charged side chains Asp6 and Arg11 and the charged amino-terminal group $NH_3^+$. Figure 14.4 also depicts surface grooves or ridges at different orientations.

Figure 14.5A shows the backbone and clustering of side chains for the cyclic ring of compstatin between residues 2 and 12, using the NMR structure.[21] Interestingly, the side chain of Trp7 bends over and caps the β-turn, in a nearly orthogonal orientation to the backbone plane. The orientation of the Trp7 side chain shows directional preference for the phenyl ring towards the side chain of Val4, although at distances greater than 6 Å from Val4, and directional preference for the indole ring towards the solvent. Other main features of the structure are the orientation of Val4 and His9 towards the interior of the cyclic ring of compstatin, the extension of the side chains of Gln5 and Asp6 to opposite directions, the clustering of Val3 with Cys2-Cys12, and the disorder of His10 and Arg11 (Figure 14.5A). Figure 14.5B

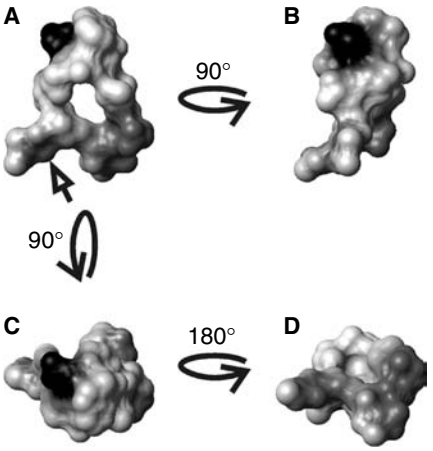


**FIGURE 14.4** Molecular (contact) surface representation with mapped electrostatic potential of the lowest energy structure of compstatin from NMR data.[21] The structural orientations in the various panels are related by 90°- or 180°-rotations as shown. In (A) and (B), the orientations of compstatin are as in Figure 14.3A and B. (C) Top view of the β-turn. (D) Bottom view of the disulfide bridge. The location of the amino-terminal $NH_3^+$ group (see discussion in text) is shown by an arrow. Black represents negative potential, dark grey represents positive potential, and light grey represents zero potential. (Coordinates from PDB code 1A1P.)

through D shows the relative topology of groups of side chains in the absence of the backbone.

Figure 14.6 shows the backbone of the best-defined structural segment of compstatin Val3-Val4-Gln5-Asp6-Trp7-Gly8-His9. This segment is best defined because of low root mean square deviation (rmsd) for backbone and side-chain atoms (Figure 14.5A–D), owed to the observation of long-range NOEs involving side chains[21] (discussed in text and supplementary material of Reference 21). The rmsd is a measure of the precision of the NMR structure. The long-range NOEs are responsible for the NMR-based "computational folding" of the peptide. The remaining part, Ile1-Cys2/His10-Arg11-Cys12-Thr13-$NH_2$, is less structured and possibly more flexible because long-range NOEs were not observed in the NMR data.

Subsequent studies of molecular dynamics simulations demonstrated the presence of several interconverting conformers.[38] The family of 21 NMR structures, the average minimized NMR structure, and the global optimization structure were used as input structures for the simulations. Twenty-three molecular dynamics trajectories were calculated for a total of 1 ns of simulation time. At 1 ns, an ensemble of five families of conformers were identified with variable populations (Figure 14.1B–F). These were: (a) coil with type I β-turn (Figure 14.1B and 14.2B), 43.5% of population; (b) β-hairpin with type I β-turn, 17.4% of population (Figure 14.1C and 14.2B); (c) β-hairpin with type II′ β-turn, 21.7% of population (Figure 14.1D and 14.2C); (d) β-hairpin with type VIII β-turn, 8.7% of population (Figure 14.1E and 14.2D); and (e) coil with α-helix, 8.7% of population (Figure 14.1F).[38]

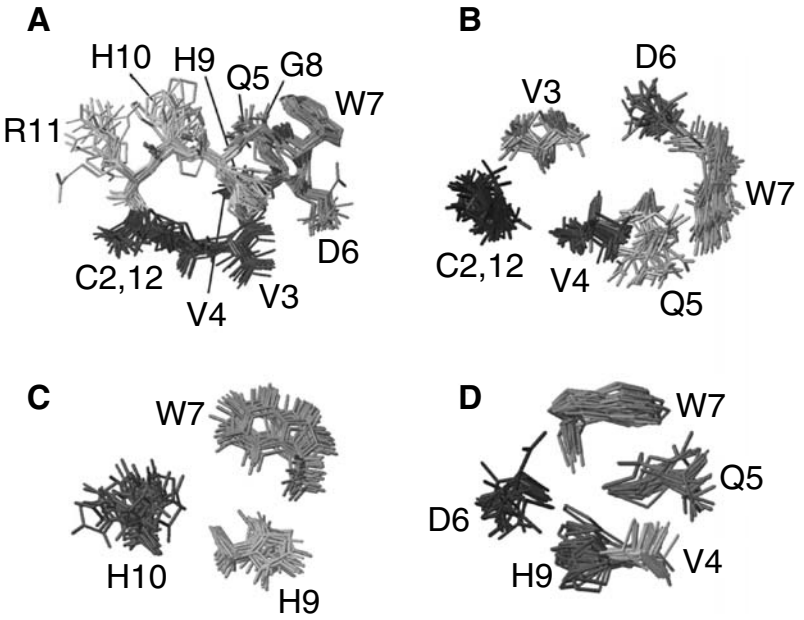


**FIGURE 14.5** Relative topology and structural precision of important clusters of side chains, using the family of 21 nuclear magnetic resonance structures of compstatin.[21] (A) Side chains within the cyclic ring comprise residues 2 to 12. Residues Ile1, Thr13, and $NH_2$ blocking group have been deleted for clarity. Hydrogen atoms have also been deleted for clarity. The backbone is shown to facilitate tracing the sequence and structure. The β-turn, Gln5-Asp6-Trp7-Gly8, is drawn in grey; the hydrophobic cluster, Cys12-Cys2-Val3-Val4, is drawn in dark grey; and the remaining polar segment, His9-His10-Arg11 is drawn in light grey. (B) Side chains of the hydrophobic cluster and β-turn of compstatin. Hydrophobic cluster residues within the cyclic ring Cys12-Cys2-Val3-Val4 and β-turn residues Gln5-Asp6-Trp7 are drawn. Hydrogen atoms are shown. (C) Side chains of ring-containing residues Trp7, His9, and His10. Hydrogen atoms are shown. (D) Side chains of β-turn residues Gln5-Asp6-Trp7 and flanking residues Val4, His9. The turn-flanking positions have been optimized to yield analogues with higher activity, using rational and experimental and computational combinatorial design. Hydrogen atoms have been deleted for clarity. Different but consistent tones of grey have been used for each residue in panels (B) through (D). (Coordinates from PDB code 1A1P.)

In both studies, NMR and molecular dynamics, the structure of the major conformer of compstatin was coil with type I β-turn. The population of the major conformer of compstatin was estimated to be 42% to 63% by the NMR data, using $^3J_{HN\text{-}H\alpha}$ coupling constants of residues 2 and 3 (Asp6 and Trp7) of the type I β-turn.[21] This is in agreement with the population of the major conformer of the molecular dynamics simulations, which was 43.5%.[38]

The molecular dynamics data provided additional quantitative measurement of the conformational interconversion of compstatin.[38] The conformational switch was possible with small amplitude motions of the backbone atoms in the range of 0.1 to 0.4 Å. These motions demonstrate the spatial similarity of the identified distinct conformers. Also, the conformational switch involved crossing of free energy barriers

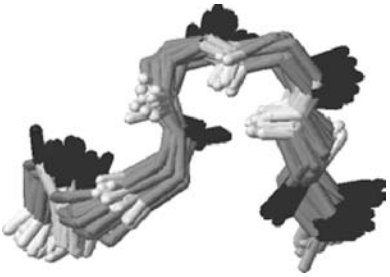

**FIGURE 14.6** Superimposition of the family of 21 nuclear magnetic resonance structures,[21] using only the best-defined segment. This segment comprises residues 3 to 9 and was assessed as "best defined" because of the presence of long-range nuclear Overhauser effects. The following color atom code has been used: black for C′ = O bond, grey for $C^{\alpha}$–C′ and $C^{\alpha}$–N bonds, and white for N–H bond. (Coordinates from PDB code 1A1P.)

in the range of 2 to 11 kcal/mol. The calculated gain or loss of free energy corresponds to the formation or deformation of one to six hydrogen bonds, associated with structure formation and deformation. These hydrogen bonds could be interstrand stabilizing β-hairpins, intrahelical stabilizing turns of α-helices, or other hydrogen bonds involving side chains. Compensatory effects involving backbone–backbone, side chain–backbone, or side chain–side chain hydrogen bonds may be present contributing to the overall free energy value.

## VI. SEQUENCE–STRUCTURE–ACTIVITY CORRELATIONS

Table 14.1 shows that the flanking residues of the ring structure contribute to increased activity by around two- to threefold.[21] This finding demonstrates that although the activity of compstatin is owed to the sequence and structure within the 11-membered ring, the 13-membered peptide is amenable to optimization.

An alanine scan was performed for every residue within the cyclic loop of the analogue that lacked the loop-flanking terminal residues, [Cys-Val-Val-Gln-Asp-Trp-Gly-His-His-Arg-Cys]-$NH_2$.[21] Also, Cys2 and Cys12 were replaced by Ala to form an inactive linear analogue with ring-flanking residues.[22] These studies showed replacement of His9 by Ala yielded about a twofold more active analogue, while replacements of Val4, His10, and Arg11 by Ala yielded about twofold less active analogues. Finally, replacements of Cys2, Val3, Gln5, Gly8, and Cys12 by Ala yielded analogues that were inactive (an experimental limit of 18-fold loss or more was set to define inactivity using alternative pathway assays).[21,22] Also, replacement of Asp6 and Trp7 by Ala resulted in lower activity, yielding about eightfold and sixfold less active analogues.[21] A similar study with the 13-membered unblocked compstatin confirmed these results with the exception of Trp7Ala replacement, which yielded an inactive analogue, and Asp6Ala replacement, which yielded an approximately fivefold less active analogue.[27] The small differences may be related to

different cutoff values defining inactivity, or differences in peptide length and terminal blocking, or assay differences in the two studies.

Interestingly, there was a correlation between the residues that resulted in loss or significant reduction of activity, and their location in the part of the structure of compstatin that was important for structural stability, the disulfide bridge/hydrophobic cluster and β-turn. This was the first sequence–structure–activity correlation that formed the basis for optimization of the sequence of compstatin.[21,23] Based on the structural stability and sequence–activity data, it was judged that residues Cys12-Cys2-Val3 of the hydrophobic cluster and Gln5-Asp6-Trp7-Gly8 of the β-turn were indispensable for high inhibitory activity. Figures 14.3 and 14.5 show the spatial clustering of the two sets of residues in the NMR structures. It was hypothesized that the hydrophobic cluster and the β-turn were important for binding to C3. The remaining six residues — Ile1, Val4, His9, His10, Arg11, and Thr13 — were thought to be amenable to further optimization.[15,17,21,23]

## VII. FIRST GENERATION OF RATIONAL DESIGN OF COMPSTATIN ANALOGUES

A major breakthrough at the time was the discovery that acetylation at the amino-terminus produced a threefold increase in activity (Table 14.1).[22,27] This analogue was called Ac-compstatin.[22] Originally, acetylation was used to block limited slow proteolytic cleavage of Ile1, which has a backbone $NH_3^+$ group.[22] Later it was shown that compstatin and Ac-compstatin maintained their integrity in diluted human serum within 5% of peptide quantity in an experiment that lasted 25 minutes.[25] This result was expected, considering the 0.03% per minute (37°C) biotransformation rate at Ile1 in human blood, for Ac-compstatin.[22] Yet inhibition experiments that were conducted within 25 minutes from adding compstatin in hemolytic assays, showed that Ac-compstatin was threefold more active than compstatin. This means that an additional mechanism is the main contributor to the higher activity produced by acetylation.

Interpretation of the mechanism for the increased activity of Ac-compstatin was based on the structure of compstatin. The positive charge of the unblocked amino-terminus was disruptive of the hydrophobic clustering at the linked termini, which was thought to participate in binding to C3 through favorable hydrophobic interactions. It was hypothesized that removal of the positive charge by blocking with the acetyl group strengthened the hydrophobic clustering and the interaction with C3, thus increasing the activity.[23] To prove the charge elimination hypothesis two analogues were designed, which reincorporated charge at the side chain now of residue 1 in Ac-compstatin, imitating the presence of charge in the unblocked peptide. In one analogue, Ile1 was replaced by Arg introducing a positive charge to disrupt the hydrophobicity of the termini, and in the other analogue Ile1 was replaced by negatively charged Asp.[25] The Ac-Ile1Arg analogue showed about a twofold reduction of inhibitory activity (Table 14.1) in agreement with the about threefold increase observed upon acetylation; the small difference is owed to the spatial location of the charge at the side chain instead of the backbone. The Ac-Ile1Asp analogue

showed an approximate fivefold reduction in inhibitory activity (Table 14.1), suggesting that a negative charge is less favorable than a positive charge in this region.

A retro–inverso analogue of compstatin that was resistant to proteolytic cleavage was tested and found inactive[22] In a retro–inverso analogue, all natural L-amino acids are replaced by their corresponding D-amino acid isomers, which are resistant to proteolysis, and the order of the sequence is reversed. The side chains retain the orientation of the L-amino acids but the backbone amide and carbonyl groups are reversed.[39] The inactive compstatin analogue had the sequence [DCys-DVal-DVal-DGln-DAsp-DTrp-DGly-DHis-DHis-DArg-DCys]. This observation supported the hypothesis that the β-turn part of the backbone was essential for activity.[22] The significance of the β-turn was further tested. Analogues that were expected, because of their propensities, to alter the β-turn type[22] or were shown by NMR to alter or abolish the β-turn[23] were inactive. Analogues with D-amino acids within the β-turn were also tested,[22,27] but they were inactive, with exception of Gly8DAla replacement, which showed some activity.[27]

## VIII. SECOND GENERATION OF RATIONAL DESIGN OF COMPSTATIN ANALOGUES

In our efforts to identify the contribution of individual compstatin residues to structural specificity and stability we designed a number of analogues that aimed to perturb the previously identified key structural elements. These were the disulfide bridge, the β-turn, and the hydrophobic cluster.[17,22,23] The structural perturbations aimed to enhance, disrupt, or alter the structure and dynamic character of compstatin.[23] The designed analogues were studied by 2D NMR spectroscopy, their spectra were compared to the spectra of compstatin[21] and Ac-compstatin,[25] and conclusions on structural similarities or differences were reached. The NMR data together with binding and inhibitory activity data allowed for structure–dynamics-binding–activity correlations.[23]

Seven analogues were studied with substitutions that introduced the following perturbations:[23] (a) linearized the peptide, (b) disrupted the hydrophobic clustering, (c) introduced flexibility inside the β-turn, (d) introduced flexibility immediately after the carboxy-terminal end of the β-turn, (e) introduced flexibility outside both ends of the β-turn, (f) attempted to switch from type I to type II β-turn, and (g) probed the role of the side chain of Trp7 in structure and activity. The NMR and activity studies,[23] in combination with previous results from the alanine scan,[21] the first round of rational design,[22] and the structures of compstatin[21,37] and Ac-compstatin[25] led to the following conclusions:

1. The linear analogue showed that the sequence of compstatin has propensity for structure formation consistent with a turn of a $3_{10}$-helix or a β-turn in the segment Gln5-Asp6-Trp7-Gly8, even in the absence of the disulfide bridge.
2. The β-turn reverses the direction of the structure and the disulfide bridge prevents the termini from drifting apart. In combination the β-turn and

the disulfide bridge introduce an optimum separation between the two arms of compstatin and aid in the formation of the hydrophobic cluster.
3. The hydrophobic cluster at the linked termini is involved in binding to C3 and in activity, but alone is not sufficient for activity.
4. The type I β-turn is a necessary but not a sufficient condition for inhibitory activity.
5. Substitutions immediately outside the two ends of the β-turn altered the turn population but not the turn structure.
6. Flexibility of the β-turn contributes to activity.
7. The sequence of β-turn residues Gln5-Asp6-Trp7(Phe7)-Gly8 is specific for turn formation but only the sequence Gln5-Asp6-Trp7-Gly8 is specific for activity.
8. Trp7 is likely to be involved in direct interaction to C3 but not of hydrophobic type. It was speculated that Trp7 may be participating in compstatin-C3 binding as a possible hydrogen bond donor through its indole amide. However, subsequent studies showed that even if this was the case, additional contributions may be present involving π–π interactions owed to aromatic ring stacking and/or π–cation interactions (see below).

Two of the seven analogues studied by NMR showed higher inhibitory activity than compstatin by factors of about three- and fourfold (analogues 15 and 17, respectively) (Table 14.1).[23] One of them, Ac-V4A/H9A/T13I, has a more flexible β-turn because of the presence of flanking Ala residues, and enhanced hydrophobic cluster because of Thr13Ile replacement. The simplicity of Ala side chains allows for more backbone conformational freedom; and Ile is more hydrophobic than Thr, which has mixed hydrophobic and polar character because of its methyl and hydroxyl groups, respectively. The other analogue, Ac-H9A, also had a slightly higher degree of flexibility than compstatin. Both of these higher activity analogues demonstrated that sequence positions 4 and 9 were amenable to further optimization. This was a major breakthrough, as evidenced by the successful subsequent experimental and computational combinatorial designs[25,40,41] (see below).

Fine-tuning of the design was performed by introducing conservative replacements in the sequence of compstatin at positions 1, 3, 5, 6, 11, and 13.[23] Activities of the conservative replacement analogues were measured but without parallel NMR studies, as we reasoned that the structures of these analogues would not deviate much from the structure of compstatin. All of the conservative replacement analogues were active, with two of them showing higher activity than compstatin. These were Ac-Q5N and Ac-T13I with about threefold and fourfold higher inhibitory activity (Table 14.1, analogues 14 and 16, respectively). Overall, the fine-tuning showed that Val is slightly preferred than Leu at position 3 (Table 14.1, analogue 9); Asn is equally preferred as Gln at position 5 (Table 14.1, analogue 14); Arg is preferred than Lys at position 11 (Table 14.1, analogue 6); and Ile is preferred than Thr at position 13 (Table 14.1, analogue 16). (Note that these arguments are made from comparisons with Ac-compstatin, analogue 13.) Special attention was paid to Arg11 replacements to address the effect on activity of proteolytic cleavage during biotransformation of compstatin. It was shown that Arg was the most preferred amino acid

at position 11, although analogues with Arg11Ser and double His9Ala/Arg11Ala replacements were also active (Table 14.1, analogues 4 and 10, respectively). One of the very first analogues of ring-only compstatin with Arg11Ala replacement was also active;[20] but, an analogue with D-Arg11 instead of Arg11 was inactive.[23] Examination of the activities of analogues with double or more replacements (Table 14.1) suggested contribution of compensatory effects and fine pairwise interactions among side chains in activity.[23]

Based on the NMR and inhibitory activity studies a sequence template was constructed of the type Ac-Xaa-[Cys-Val-Xaa-Gln-Asp-Trp-Gly-Xaa-Xaa-Xaa-Cys]-Xaa-$NH_2$, where the six amino acids named Xaa were deemed amenable to further optimization and the remaining seven amino acids were deemed indispensable for activity.[17] This sequence template, called active sequence template hereafter, was used in subsequent experimental combinatorial design[25] (see below) and together with the NMR structural template was used for subsequent computational combinatorial design[40,41] (see below).

## IX. SECOND GENERATION OF EXPERIMENTAL COMBINATORIAL DESIGN

A second round of phage-displayed design was performed[25] using the active sequence template Ac-Xaa-[Cys-Val-Xaa-Gln-Asp-Trp-Gly-Xaa-Xaa-Xaa-Cys]-Thr-$NH_2$, which was derived from the rational design. The use of phage-displayed random peptide libraries has also been called experimental combinatorial design[15] to distinguish it from computational combinatorial design (see below). Only position Xaa was randomized, where Xaa represents any amino acid. The combinatorial gene sequences NNS-TGC-GTG-NNS-CAG-GAC-TGG-GGC-$(NNS)_3$-TGC-NNS, were displayed at the amino-terminus of the phage.[25] In these sequences, N represents any of the four nucleotides, A, C, G, T, and S represents C, G, T, in equal molar ratios.

This search resulted in four active clones.[25] Synthetic peptides with the sequence of the active clones were tested for activity and showed that one of them was about half as active as compstatin and the other about fourfold more active than compstatin (Table 14.1, analogues 3 and 18, respectively). The innovation in these two analogues is the incorporation of side chains with aromatic rings at positions 4 and 9, such as Phe4 (Table 14.1, analogue 3) and Trp9 (Table 14.1, analogue 18). The remaining two analogues showed His and Asp at position 4, but their activity was lower and much lower, respectively, compared to analogues 3 and 8 of Table 14.1. We have speculated that electronic effects owed to the relative orientation (stacking) of the ring of Trp7 and Trp9 (Table 14.1, analogue 18) or Phe4 and Trp7 (Table 14.1, analogue 3) were contributors to activity. The electronic effects may be of the type of π–π dipole–dipole or dipole-induced dipole polarization interactions or π-cation or π–π-cation interactions with a possible cation (in Arg, Lys, or charged His side chains) located on the C3-binding site. The possibility of structure stabilization with π-cation interactions within compstatin, involving Arg11 or one of the histidines, was also considered; but this was not supported by the NMR or MD structures of compstatin.[49] In addition, theoretical calculations of apparent $pK_a$ values of ionizable

residues showed that His9 and His10 were predominantly neutral at physiological pH (D. Morikis and J.D. Lambris, unpublished data), which excludes a contributing cation from their side chains.

NMR studies of the most active analogue of the second round of phage-display design, Ac-I1L/H9A/T13G (Table 14.1, analogue 18), showed that its structure was consistent with the structure of Ac-compstatin.[25] This observation excludes structural changes from being responsible for the increased activity of the analogue. It should be noted that both the rational design and experimental combinatorial design yielded fourfold more active analogues than compstatin (Table 14.1, analogues 17 and 18, respectively), involving His9Ala and His9Trp replacements. Since Ala and Trp have very different side chains, it is possible that they contribute to activity through different mechanisms. For example, Ala9 in the Ac-His9Ala analogues introduces flexibility that facilitates binding; and the orientation of Trp9 relative to Trp 7 in the Ac-I1L/H9A/T13G analogue contributes to favorable electronic interaction of the rings of the two tryptophans that promotes better binding. In addition to possible involvement of the aromatic rings of Trp residues in binding and inhibitory activity, hydrogen bond formation involving either or both Trp7 and Trp9 as donors was also considered as possible.

## X. FIRST GENERATION OF COMPUTATIONAL COMBINATORIAL DESIGN

Compstatin was used as the first test case[40–42] for a novel computational combinatorial methodology for drug design developed by the Floudas group.[37,43–47] The methodology involved two steps, one at the sequence selection level and another at the structure validation level. The first step was based on a mixed-integer linear optimization algorithm that used a distance-dependent backbone potential with implicit inclusion of side chain interactions and specificities.[37,48] The algorithm selected and ranked several possible sequences that were compatible with a structural template that provides the Cα–Cα inter-atomic backbone distances. The second step was based on a global optimization algorithm that used a full-atom force field.[43,44] The algorithm calculated ensemble probabilities for the selected sequences applied on flexible structural templates.[37] In the first step the active sequence template of compstatin, Ac-Xaa-[Cys-Val-Xaa-Gln-Asp-Trp-Gly-Xaa-Xaa-Xaa-Cys]-Thr-$NH_2$,[23,25,40,41] which was identified by rational design, was used (see above). This is the same sequence template that was used in the second round of experimental combinatorial design (see above). Also, in both first and second steps, structural templates from the NMR-derived structures of compstatin were used. The averaged NMR structure provided the structural template in the first step and the family of NMR structures provided the flexible templates of the second step, with flexibility being determined by the structural variation of the family of NMR structures. The underlying assumption of this approach was that amino acid specificity (first step) and the predicted increase in fold stability and specificity (second step), were correlated with increase in functionality, while structural characteristics essential for function were maintained (in active sequence and structural templates).

The active sequence template of compstatin that allowed optimization of 6 out of 13 residues was used to reduce the combinatorial challenge of the problem. The number of possible combinations becomes astronomic as the number of residues is increased in systematic computational combinatorial optimization. This is not so much of an issue in experimental combinatorial design, where the phage-displayed libraries are constructed in a random manner, which resembles an evolutionary optimization. For example, if we allow that the number of combinations for the 20 amino acids in a dipeptide is $20^2 = 400$. For a tripeptide, the number is $20^3 = 8000$; for a six-residue peptide, $20^6 = 64{,}000{,}000$; and for a 13-residue peptide, $20^{13} = 8.2 \times 10^{16}$. To further reduce the combinatorial challenge, small amino acid groups, formed by common physicochemical properties, were used in the calculations for the six positions amenable to optimization.[40,41] The hydrophobic amino acid set (Ala, Phe, Ile, Leu, Met, Val, Tyr) was used for positions 1, 4, and 13. Partially hydrophobic Thr was also added to the hydrophobic set for position 13 to account for the parent peptide residue at this position. With the exception of Cys and Trp, all remaining 18 amino acids were allowed for positions 9, 10, and 11.

The results were impressive, predicting several active analogues.[40,41] Five of these analogues were found to be 6- to 16-fold more active than compstatin (Table 14.1, analogues 19–23). A common characteristic of these five analogues is the presence of Tyr at position 4, a first-time finding. Another finding is the presence of Phe at position 9 in two of these analogues, while the remaining three maintained Ala at this position. It is possible that the pair Tyr4-Trp7 or the triplet Tyr4-Trp7-Phe9 play key roles in the activity of compstatin.

The computational combinatorial design results (Table 14.1)[40,41] were consistent with the findings of the experimental combinatorial design that also identified aromatic ring residues at positions 4 and 9 (Table 14.1).[25] The most active analogue identified by the computational combinatorial methodology is Ac-V4Y/H9A, which combines the replacements of Val4Tyr and His9Ala (Table 14.1, analogue 23).

## XI. DESIGN OF MOST ACTIVE PEPTIDE ANALOGUE

The successes of the three approaches — rational, experimental combinatorial, and computational combinatorial design — used in the optimization of compstatin prompted a critical examination of the replacements at positions 4 and 9 that showed increased activity upon replacement with residues containing aromatic rings. When an analogue was designed with Trp at position 4 and Ala at position 9, it showed a ~45-fold higher activity than compstatin.[49] The Ac-V4W/H9A is currently the peptide analogue of compstatin composed of all-natural amino acids with the highest inhibitory activity (Table 14.1, analogue 24). It should be noted that Trp was not included in the first round of computational combinatorial design for technical reasons, but was also predicted in a subsequent round of computational combinatorial design.[42] Structural studies by NMR and structural and dynamics studies by molecular dynamics simulations have been performed to explore structure–dynamics–activity correlations.[49] A molecular dynamics simulation has shown the presence of aromatic ring interaction between Trp4 and Trp7 in the most active analog with natural amino acids, Ac-V4W/H9A (see Table 14.1, analogue 24).[49]

The identification of aromatic residues at positions 4 and 9, in addition to Trp at position 7, in analogues with higher inhibitory activity has led us to design a series of hybrid peptide–peptidomimetic analogues containing non-natural amino acids. Non-natural amino acids were incorporated to test our hypotheses for the importance of possible aromatic ring interactions or aromatic ring cation interactions for structural stability, binding, or activity, or the involvement of Trp or Tyr residues in hydrogen bond formation. Several of these analogues have shown high inhibitory activity.[49] Currently, the most active analogue with non-natural amino acids shows 99-fold higher activity than compstatin.[49] This analogue, Ac-V4(2 Nal)/H9A, has 2-naphthylalanine at position 4 and alanine at position 9. The work with non-natural amino acids is now in press[49] and will be reviewed elsewhere.

## XII. BINDING KINETICS AND STRUCTURE-BINDING RELATIONS

The technique of surface plasmon resonance (SPR) was used to determine binding rates, binding constants, and relative affinities, using BIACORE technologies for data collection and analysis. Kinetic measurements were a second step to study binding in a quantitative way, once identification of inhibition and binding targets (C3, C3b, C3c) was made using immunological assays. Three binding kinetics studies were performed.[22,25,26]

In the first study, the analogue Ac-Ile-[Cys-Val-Val-Gln-Asp-Trp-Gly-His-His-Arg-Cys]-Thr-Ala-Gly-His-Met-Ala-Asn-Leu-Thr-Ser-His-Ala-Ser-Ala-Lys-biotin, immobilized on the sensor chip through biotin, was chosen for binding kinetic measurements.[22,25] The 13 amino-terminal residues of this peptide corresponded to Ac-compstatin, and the spacer peptide corresponded to the original sequence identified by the phage-displayed random peptide library. It was reasoned that the spacer would help increase the accessibility of the active amino-terminal 13-residue peptide segment to its target. This study showed that the selected compstatin analogue bound to human C3, hydrolyzed C3($H_2O$), C3b, and C3c,[22] in agreement with previous ELISA studies.[20] Another peptide with sequence biotin-Lys-Tyr-Ser-Ser-Ile-[Cys-Val-Val-Gln-Asp-Trp-Gly-His-His-Arg-Cys]-Thr-$NH_2$, immobilized on the sensor chip through biotin linked to the amino-terminus, failed to bind C3. Compstatin was oriented through its carboxy-terminus in the phage when it was first discovered using the phage-displayed random peptide library. These observations demonstrated that a free amino-terminus is important for binding.[22]

In the second study, the binding kinetics against baboon C3 were measured for the peptide Ac-Ile-[Cys-Val-Val-Gln-Asp-Trp-Gly-His-His-Arg-Cys]-Thr-Ala-Gly-His-Met-Ala-Asn-Leu-Thr-Ser-His-Ala-Ser-Ala-Lys-biotin, which corresponded to Ac-compstatin.[26] This study also showed lack of binding to human C4 and C5, which are homologues of C3, and mouse and rat C3.

In the third study, the binding against C3 was studied for the following three analogues:[25] (a) Ac-Ile-[Cys-Val-Val-Gln-Asp-Trp-Gly-His-His-Arg-Cys]-Thr-Ala-Gly-His-Met-Ala-Asn-Leu-Thr-Ser-His-Ala-Ser-Ala-Lys-biotin, corresponding to Ac-compstatin; (b) Ac-Ile-[Cys-Val-Val-Gln-Asp-Trp-Gly-Ala-His-Arg-Cys]-Thr-Ala-Gly-His-Met-Ala-Asn-Leu-Thr-Ser-His-Ala-Ser-Ala-Lys-biotin, corresponding

to Ac-H9A (Table 14.1, analogue 17); and (c) Ac-Leu-[Cys-Val-Val-Gln-Asp-Trp-Gly-Trp-His-Arg-Cys]-Gly-Ala-Gly-His-Met-Ala-Asn-Leu-Thr-Ser-His-Ala-Ser-Ala-Lys-biotin, corresponding to Ac-I1L/H9W/T13G (Table 14.1, analogue 18).

A variety of interaction patterns were observed for each of the analogues and for the different targets and species, suggesting complexity of the binding models.[22,25,26] We will not present here detailed kinetic models, derived from the BIACORE data fitting and analysis software, because they need to be cross-validated using another method. One of the limitations of SPR is the need to attach one of the reactants in a membrane cell to expose the flow of the other reactant over it. This experimental arrangement restricts the diffusional movement of one of the reactants. In addition, the evaluation of $k_{off}$ rate is not direct. Given the limitations of SPR, it was deemed necessary to study the kinetics and thermodynamics of binding using an independent method such as isothermal titration calorimetry (ITC). Both SPR and ITC measure binding against a unique target in a single reaction, without the need for immunological assays. Immunological assays contain several species and are prone to several possible reactions and complex binding-interaction-activation schemes. In addition, ITC is a direct method to study binding in solution phase without the need to attach and partially immobilize one of the reactants in membranes. There is an upper limit for the measurement of binding constant ($\sim 10^9$ $M^{-1}$) using ITC. Compstatin is amenable to ITC study because its binding constant is below the upper limit. The ITC studies of compstatin binding to C3 are now in press.[50]

Despite the limitations of the kinetic data, a working hypothesis was constructed for the binding of compstatin to C3, based on structure–binding correlations.[22,23] The binding site of compstatin on C3 has been broadly localized within the 40-kD carboxy-terminal half of the β-chain of C3 (A.M. Soulika and J.D. Lambris, unpublished data). Figure 14.7 shows cartoon block diagrams of the targets of compstatin binding, C3, C3b, and C3c, and the potential binding site. It was suggested that a conformational change on C3 could be possible to facilitate the binding of compstatin.[22] A working binding model[23] suggested that recognition and binding of compstatin to C3 involves (a) interactions of the hydrophobic cluster of compstatin (residues Ile1-Cys2-Val3-Val4/Cys12-Thr13) with a partially hydrophobic binding site in C3; (b) shape complementarity that allows fit of the β-turn of compstatin within the binding site; and (c) interactions of the indole and/or the aromatic ring of Trp7 with suitable residues on C3. The latter could be hydrogen bond formation (indole ring), π stacking or π-cation interactions (aromatic ring). Similar interactions may be possible for Trp4 of the most active peptide analogue (Table 14.1), or other residues with aromatic rings at positions 4 or 9).

## XIII. CURRENT AND FUTURE DIRECTIONS IN OPTIMIZATION OF COMPSTATIN

Most of the success in the optimization of compstatin was based on structure-inhibitory activity correlations, which made available the active sequence template used in subsequent rational, experimental combinatorial, and computational

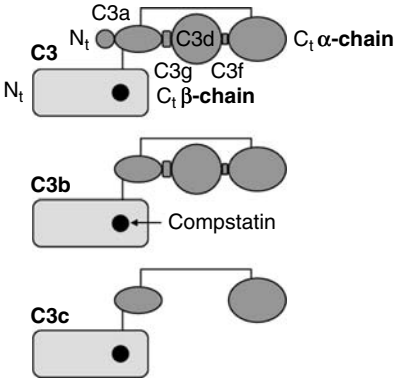


**FIGURE 14.7** Cartoon representing C3[13] with the potential binding site of compstatin. The various fragments of C3 are shown as blocks, and interchain and interfragment disulfide bridges are shown with lines. The approximate relative sizes (not in scale) of C3 components are shown with variably sized blocks. The α-chain of C3 is shown in dark grey and the β-chain is shown in light grey. The binding site of compstatin has been located within the carboxy-terminal half of the β-chain and is represented by a black circle. (A) C3. (B) C3b, produced after cleavage of C3a from C3. (C) C3c, produced after cleavage of C3d, C3f, and C3g from C3b.

combinatorial design. Key roles in the progression of discovery were played by the NMR data and their implications for structure. Determination of the structure of bound compstatin is currently in progress (D. Morikis and J.D. Lambris, unpublished data), and is expected to provide valuable input in the C3-compstatin binding. This study entails the determination of bound compstatin, but not of C3, using transfer NOE NMR data. It is expected that comparison of free and bound compstatin structures will aid to decompose the processes of recognition and binding according to the model A + B ↔ AB ↔ AB*, where A is C3 and B and B* are free and bound compstatin, respectively. The first step corresponds to recognition and the second step corresponds to binding.

A feature that has not yet been exploited is the dynamic character of compstatin as indicated by the molecular dynamics simulation data. Molecular dynamics studies of several analogues of compstatin have been used to identify the possible aromatic ring and other side chain pairwise interactions (B. Mallik and D. Morikis, unpublished data). Quantitation of pairwise interactions is expected to promote our understanding for the role of individual amino acid replacements in structural stability and structure specificity. Thus far molecular dynamics studies were performed using implicit solvent representation models. Simulations using solvent molecules explicitly have been planned and are expected to provide more accurate representation of compstatin-solvent interaction and the involved dynamics. Molecular dynamics have also provided us with multiple structural templates[38] (B. Mallik and D. Morikis, unpublished data), which will be used as quasidynamic structural templates to design a dynamic pharmacophore model. This pharmacophore model will be used for the identification of peptidomimetic or nonpeptidic C3 inhibitors. Even in the absence

of a pharmacophore model, several active hybrid peptide-peptidomimetic analogues have been identified with rational design.[49]

Accompanying data for compstatin and active analogues using isothermal titration calorimetry have been useful to explore the thermodynamics and kinetics of binding.[50] Also, cross-linking studies with photoactive labels for identification of the C3-binding site of compstatin are now in progress. Location of the C3-binding site will allow the expression of mouse C3 with a patch of human C3, comprising the C3-binding site for compstatin. The chimeric C3 will be used in animal model studies, using transgenic mice or transgenic mice expressing chimeric C3. These animal models are expected to bypass the hurdle of lack of binding of compstatin to mouse C3, and to eliminate the need of primates for routine *in vivo* studies. If the binding site on C3 is known and the relevant C3 fragment is folded, active, and of molecular mass suitable for NMR studies, we will pursue its three-dimensional structure determination. If the molecular mass of the fragment is not optimum for NMR studies co-crystallization will be attempted for structure determination by x-ray crystallography. Knowing the active site structure will open new possibilities for docking and quantitative structure–activity relations (QSAR) to study the mechanism of binding and activity, and to identify smaller nonpeptidic active molecules. The ultimate goal is to perform *in vivo* studies in primates for complement inhibitory activity and toxicity for the most potent peptide, hybrid peptide-peptidomimetic, or nonpeptidic analogues of compstatin.

## ACKNOWLEDGMENTS

We are grateful to several collaborators who have participated in various stages of this study, and all group members who have made contribution through casual discussions and academic interest. In particular, we would like to mention people who worked in several aspects of the study: Arvind Sahu, Athena Soulika, Lynn Spruce, Christodoulos Floudas, Henry Edmunds, Bo Nilsson, Kristina Nilsson-Ekdahl, Tom Molnes, Arnt Fiane, John Klepeis, William Moore, Buddhadeb Mallik, Melinda Roy, Christos Tsokos, George Tsokos, Emilia Argyropoulos, Caterina Carafides, Anastasios Troganis, Maria Rosa Sarrias, Brian Kay, and Patricia Jennings. We are also grateful for funding from the National Institutes of Health, the Walter Reed Army Institute of Research, and the American Heart Association.

## SUPPLEMENTARY MATERIAL ON CD

All figures, including Figures 14.1 through 14.6 in color, and their corresponding captions are supplied on the companion CD. Coordinates for the average NMR, global optimization, and molecular dynamics structures of compstatin (not deposited at the Protein Data Bank) are included on the companion CD. (Courtesy of D. Morikis, J.D. Lambris, J.L. Klepeis, C.A. Floudas, and B. Mallik.)

## REFERENCES


1. MJ Walport. *N. Engl. J. Med.*, 344:1140–1144, 2001.
2. MJ Walport. *N. Engl. J. Med.*, 344:1058–1066, 2001.
3. R Barrington, M Zhang, M Fischer, MC Carroll. *Immunol. Rev.*, 180:5–15, 2001.
4. JD Lambris, KBM Reid, JE Volanakis. *Immunol. Today*, 20:207–211, 1999.
5. D Mastellos, D Morikis, SN Isaacs, MC Holland, CW Strey, JD Lambris. *Immunol. Res.*, 27:367–385, 2003.
6. D Mastellos, JD Lambris. *Trends Immunol.*, 23:485–491, 2002.
7. VM Holers. *Clin. Immunol.* 107:140–151, 2003.
8. BP Morgan, CL Harris. *Mol. Immunol.*, 40:159–170, 2003.
9. A Sahu, JD Lambris. *Immunopharmacology*, 49:133–148, 2000.
10. JD Lambris, VM Holers, Eds. *Therapeutic Interventions in the Complement System*. Humana Press, Totowa, NJ, 2000.
11. SC Makrides. *Pharmacol. Rev.*, 50:59–87, 1998.
12. CL Harris, DA Fraser, BP Morgan. *Biochem. Soc. Trans.*, 30:1019–1026, 2002.
13. A Sahu, JD Lambris. *Immunol. Rev.*, 180:35–48, 2001.
14. A Sahu, JO Sunyer, WT Moore, MR Sarrias, AM Soulika, JD Lambris. *Immunol. Res.*, 17:109–121, 1998.
15. D Morikis, AM Soulika, B Mallik, JL Klepeis, CA Floudas, JD Lambris. *Biochem. Soc. Trans.*, 32:28–32, 2004.
16. D Morikis, A Sahu, WT Moore, JD Lambris. In J Matsukas, T Mavromoustakos, Eds. *Design*, *Structure*, *Function and Application of Compstatin*. Ios Press, Amsterdam, 1999, pp. 235–246.
17. D Morikis, JD Lambris. *Biochem. Soc. Trans.*, 30:1026–1036, 2002.
18. A Sahu, D Morikis, JD Lambris. In JD Lambris, VM Holers, Eds., *Complement Inhibitors Targeting C3*, *C4*, *and C5*. Humana Press, Totowa, NJ, 2000, pp. 75–112.
19. M Szardenings. *J. Reception Signal Trans.*, 23:307–349, 2003.
20. A Sahu, BK Kay, JD Lambris. *J. Immunol.*, 157:884–891, 1996.
21. D Morikis, N Assa-Munt, A Sahu, JD Lambris. *Protein Sci.*, 7:619–627, 1998.
22. A Sahu, AM Soulika, D Morikis, L Spruce, WT Moore, JD Lambris. *J. Immunol.*, 165:2491–2499, 2000.
23. D Morikis, M Roy, A Sahu, A Troganis, PA Jennings, GC Tsokos, JD Lambris. *J. Biol. Chem.*, 277:14942–14953, 2002.
24. B Nilsson, R Larsson, J Hong, G Elgue, KN Ekdahl, A Sahu, JD Lambris. *Blood*, 92:1661–1667, 1998.
25. AM Soulika, D Morikis, MR Sarrias, M Roy, LA Spruce, A Sahu, JD Lambris. *J. Immunol.*, 171:1881–1890, 2003, Erratum at 172:5128, 2004.
26. A Sahu, D Morikis, JD Lambris. *Mol. Immunol.*, 39:557–566, 2003.
27. ST Furlong, AS Dutta, MM Coath, JJ Gormley, SJ Hubbs, D Lloyd, RC Mauger, AM Strimpler, MA Sylvester, CW Scott, PD Edwards. *Immunopharmacology*, 48:199–212, 2000.
28. SF Altschul, W Gish, W Miller, EW Myers, DJ Lipman. *J. Mol. Biol.*, 215:403–410, 1990.
29. AM Soulika, MM Khan, T Hattori, FW Bowen, BA Richardson, CE Hack, A Sahu, LH Edmunds Jr, JD Lambris. *Clin. Immunol.*, 96:212–221, 2000.
30. S Schmidt, G Haase, E Csomor, R Lutticken, H Peltroche-Llacsahuanga. *J. Biomed. Mater. Res.* 66A:491–499, 2003.
31. SH Sacks, P Chowdhury, WD Zhou. *Curr. Opin. Immunol.*, 15:487–492, 2003.

32. AE Fiane, TE Mollnes, V Videm, T Hovig, K Hogasen, OJ Mellbye, L Spruce, WT Moore, A Sahu, JD Lambris. *Xenotransplantation*, 6:52–65, 1999.
33. AE Fiane, TE Mollnes, V Videm, T Hovig, K Hogasen, OJ Mellbye, L Spruce, WT Moore, A Sahu, JD Lambris. *Transpl. Proc.*, 31:934–935, 1999.
34. TE Mollnes, OL Brekke, M Fung, H Fure, D Christiansen, G Bergseth, V Videm, KT Lappegard, J Kohl, JD Lambris. *Blood*, 100:1869–1877, 2002.
35. A Klegeris, EA Singh, PL McGeer. *Immunology*, 106:381–388, 2002.
36. HM Berman, J Westbrook, Z Feng, G Gilliland, TN Bhat, H Weissig, IN Shindyalov, PE Bourne. *Nucleic Acids Res.*, 28:235–242, 2000.
37. JL Klepeis, CA Floudas, D Morikis, JD Lambris. *J. Comp. Chem.*, 20:1354–1370, 1999.
38. B Mallik, JD Lambris, D Morikis. *Proteins Struct. Function Genet.* 53:130–141, 2003.
39. M Chorev, M Goodman. *Trends Biotechnol.*, 13:438–445, 1995.
40. JL Klepeis, CA Floudas, D Morikis, CG Tsokos, E Argyropoulos, L Spruce, JD Lambris. *J. Am. Chem. Soc.*, 125:8422–8423, 2003.
41. JL Klepeis, CA Floudas, D Morikis, CG Tsokos, JD Lambris. *Ind. Eng. Chem. Res.*, 43:3817–3826, 2004.
42. CA Floudas, JL Klepeis, JD Lambris, D Morikis. In CA Floudas, R Agrawal, Eds., *Proceedings FOCAPD 2004*, Sixth International Conference on Foundation of Computer-Aided Process Design, Discovery through Products and Process Design, 2004.
43. JL Klepeis, CA Floudas. *J. Global Optimization*, 25:113–140, 2003.
44. JL Klepeis, HD Schafroth, KM Westerberg, CA Floudas. *Adv. Chem. Phys.*, 120:265–457, 2002.
45. JL Klepeis, CA Floudas. *J. Chem. Phys.*, 110:7491–7512, 1999.
46. CA Floudas. *Nonlinear and Mixed-Integer Optimization: Fundamentals and Applications*. Oxford University Press, Oxford, 1995.
47. CA Floudas. *Deterministic Global Optimization: Theory, Methods and Applications: Nonconvex Optimization and Its Applications*. Kluwer, Dordrecht, 2000.
48. D Tobi, R Elber. *Proteins, Struct. Function Genet.*, 41:40–46, 2000.
49. B Mallik, M Katragadda, L Spruce, C Carafides, CG Tsokos, D Morikis, JD Lambris. *J. Med. Chem.,* 48:274–286, 2005.
50. M. Katragadda, D Morikis, JD Lambris. *J. Biol. Chem.*, 279:54987–54995, 2004.
51. R Koradi, M Billeter, K Wuthrich. *J. Mol. Graph.*, 14:51–55, 1996.